

Рассмотреть оптимальную скользящую среднюю количества лидов для автобампа.

На основании исторических данных с 2020/01/01 по всем странам и продуктам.

Рассматривались распределения, отклонения, а также основные параметры мер центральной тенденции для случаев без выбросов, с усеченными выбросами и полностью без выбросов.

Исследование было для дней, недель и месяцев.

Сводная таблица и заключение в конце.

Days

```
In [164]: df_days = df
```

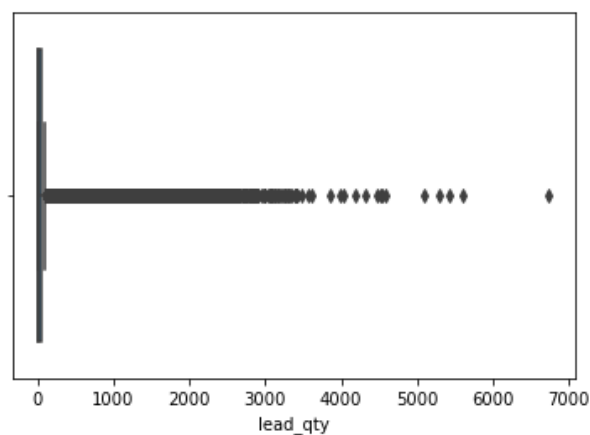
```
In [165]: df.describe()
```

Out[165]:

	lead_qty
count	240785.000000
mean	60.935557
std	179.100450
min	1.000000
25%	2.000000
50%	8.000000
75%	42.000000
max	6734.000000

```
In [166]: sns.boxplot(x='lead_qty', data=df_days)
```

Out[166]: <AxesSubplot:xlabel='lead_qty'>



lead_qty

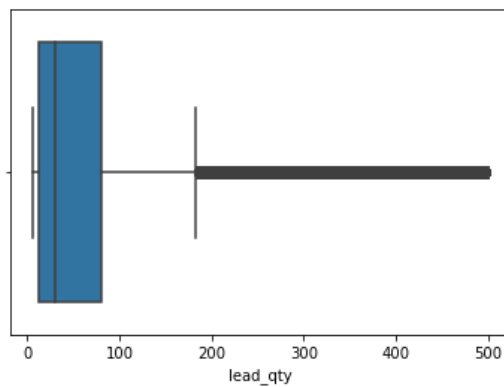
```
In [253]: df_days_cut = df_days[(df_days['lead_qty'] < 500) & (df_days['lead_qty'] > 5)]
df_days_cut.describe()
```

Out[253]:

lead_qty	
count	132756.000000
mean	67.735665
std	89.630838
min	6.000000
25%	13.000000
50%	31.000000
75%	81.000000
max	499.000000

```
In [254]: sns.boxplot(x='lead_qty', data=df_days_cut)
```

Out[254]: <AxesSubplot:xlabel='lead_qty'>



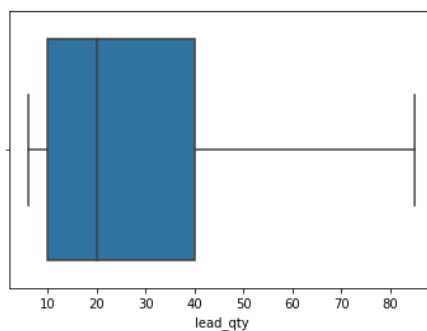
```
In [255]: df_days_fullcut = df_days[(df_days['lead_qty'] < 86) & (df_days['lead_qty'] > 5)]
df_days_fullcut.describe()
```

Out[255]:

lead_qty	
count	101101.000000
mean	27.526206
std	20.974809
min	6.000000
25%	10.000000
50%	20.000000
75%	40.000000
max	85.000000

```
In [256]: sns.boxplot(x='lead_qty', data=df_days_fullcut)
```

Out[256]: <AxesSubplot:xlabel='lead_qty'>



```
In [330]: round(compare_df)
```

```
Out[330]:
```

	days	weeks	months	-----	days_cut	weeks_cut	months_cut	-----	days_fullcut	weeks_fullcut	months_fullcut
count	240785.0	53412.0	18082.0		132756.0	20473.0	5065.0		101101.0	15120.0	3799.0
mean	61.0	275.0	811.0		68.0	447.0	1856.0		28.0	165.0	693.0
std	179.0	997.0	3495.0		90.0	610.0	2604.0		21.0	123.0	519.0
min	1.0	1.0	1.0		6.0	36.0	151.0		6.0	36.0	151.0
25%	2.0	3.0	3.0		13.0	80.0	331.0		10.0	64.0	270.0
50%	8.0	17.0	24.0		31.0	197.0	794.0		20.0	123.0	509.0
75%	42.0	131.0	230.0		81.0	527.0	2125.0		40.0	239.0	1013.0
max	6734.0	33767.0	108318.0		499.0	3498.0	14988.0		85.0	501.0	2125.0

Чтобы оценить оптимальную скользящую, я рассчитал «коэффициенты роста» стандартного отклонения для разных случаев на основании дневного значения.

```
In [345]: 179*7/997 # неделя с выбросами
```

```
Out[345]: 1.2567703109327983
```

```
In [346]: 179*30/3495 # месяц с выбросами
```

```
Out[346]: 1.536480686695279
```

```
In [ ]:
```

```
In [347]: 90*7/610 # неделя с частичными выбросами
```

```
Out[347]: 1.0327868852459017
```

```
In [349]: 90*30/2604 # месяц с частичными выбросами
```

```
Out[349]: 1.0368663594470047
```

```
In [ ]:
```

```
In [350]: 21*7/123 # неделя без выбросов
```

```
Out[350]: 1.1951219512195121
```

```
In [351]: 21*30/519 # месяц без выбросов
```

```
Out[351]: 1.2138728323699421
```

Если отбросить основные выбросы, то оптимальные показатели оказываются на недельной отсечке.

То есть, вебмастер показывает свою усредненную эффективность уже на недельном промежутке времени.

При месячном отрезке, коэффициент оказывается немного лучше, но абсолютно незначительно. При этом появляется риск, что при более эффективной работе вебмастера в течение месяца, ему будет постоянно начисляться бамп. Поэтому оптимально рассматривать эффективность вебмастера **на недельных промежутках времени**.